

USER GUIDE

APLIKASI EKSTRAKSI TABEL HTML

```

...
<table cellpadding="2" cellspacing="2" border="1" width="30%">
  <tbody>
    <tr>
      <td valign="top" align="center"><b>No<br>
      </b></td>
      <td valign="top" align="center"><b>Nama<br>
      </b></td>
      <td align="center"><b>Kode Area</b><br>
      </td>
    </tr>
    <tr>
      <td align="center">1</td>
      <td align="center">Il. Angrek</td>
      <td align="center">Jakarta</td>
    </tr>
    <tr>
      <td align="center">2</td>
      <td align="center">Il. Mawar</td>
      <td align="center">Surabaya</td>
    </tr>
    <tr>
      <td align="center">3</td>
      <td align="center">Il. Sakura</td>
      <td align="center">Semarang</td>
    </tr>
    <tr>
      <td align="center"><b>Ali<br>
      </b></td>
      <td align="center" rowspan="1" colspan="1" align="left">021<br>
      </td>
      <td align="center">$612571<br>
      </td>
    </tr>
  </tbody>
</table>

```

NO	IAIAN	KOTA
1	Il. Angrek	Jakarta
2	Il. Mawar	Surabaya
3	Il. Sakura	Semarang

Gunadarma
University
<http://www.gunadarma.ac.id>



Oleh:
Prof. Dr. I Wayan Simri Wicaksana
Dr. Detty Purnamasari
Dr. Lintang Yuniar Banowosari
Lucky Rachmadeni, ST.

Universitas Gundarma
Desember 2013

COLORING THE GLOBAL FUTURE

USER GUIDE **APLIKASI EKSTRAKSI TABEL HTML**

Oleh:
Prof. Dr. I Wayan Simri Wicaksana
Dr. Detty Purnamasari
Dr. Lintang Yuniar Banowosari
Lucky Rachmadeni, ST.

Universitas Gunadarma
Desember 2013



KATA PENGANTAR

Puji Syukur Kehadirat Tuhan Yang Maha Esa, atas berkat RidhoNya Aplikasi Ekstraksi Tabel HTML telah selesai dikembangkan dan juga telah dilengkapi oleh Buku *User Guide* aplikasi. Buku *User Guide* ini merupakan panduan yang digunakan untuk dapat menjalankan aplikasi ekstraksi tabel HTML, yang di dalamnya dijelaskan spesifikasi minimum komputer, dan langkah-langkah proses instalasi untuk kebutuhan program yang harus ada pada komputer, serta contoh penggunaan aplikasi yang juga disajikan dalam bentuk gambar, sehingga memudahkan pengguna dalam menjalankan aplikasinya.

Aplikasi ekstraksi tabel HTML merupakan suatu aplikasi yang dikembangkan untuk membantu pengguna Internet mengambil dan menggunakan kembali informasi yang disajikan dalam bentuk tabel di Internet. Tabel yang terdapat pada Internet dalam format HTML yang telah di ekstrak, disimpan dalam format database.

Kehadiran buku *User Guide* ini diharapkan dapat membantu pengguna untuk menjalankan aplikasi ekstraksi tabel HTML, dan semoga aplikasi ini bermanfaat bagi pengguna Internet khususnya, dan masyarakat luas pada umumnya, yang membutuhkan program ekstraksi tabel yang berformat HTML. Penulis mengucapkan terimakasih kepada Universitas Gunadarma Jakarta yang memberikan dukungan sehingga aplikasi dan buku *User Guide* ini dapat diselesaikan. Masukan pembaca baik untuk aplikasi dan buku *User Guide* ini merupakan hal yang membahagiakan bagi kami.

Penulis
Desember 2013



DAFTAR ISI

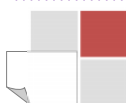
Cover	
Kata Pengantar	
Daftar Isi dan Daftar Gambar	
1. Spesifikasi Komputer	1
2. Cara Instalasi Bahasa Pemrograman Python	1
3. Cara Instalasi Xampp dan MySQL	7
4. Cara Instalasi Library Python	12
5. Cara Konfigurasi dan Menjalankan Aplikasi EKstraksi Tabel	14

DAFTAR GAMBAR

Gambar 1. File Python 2.7 Installer	1
Gambar 2. Jendela Setup "Python for windows"	2
Gambar 3. Jendela "select destination directory"	3
Gambar 4. Jendela "customize Python 2.7"	3
Gambar 5. Jendela Progress Bar Proses Instalasi	4
Gambar 6. Jendela "completing the Python 2.7 Installer"	4
Gambar 7. Jendela "properties" dan "advanced system setting"	5
Gambar 8. Jendela "system properties"	5
Gambar 9. Jendela "environment variables"	6
Gambar 10. Jendela Variable "path"	6
Gambar 11. File Xampp Installer	7
Gambar 12. Jendela "set-up Xampp"	8
Gambar 13. jendela "select component"	8
Gambar 14. jendela "installation folder"	9
Gambar 15. jendela "Bitnami for Xampp"	9
Gambar 16. jendela "ready to install"	10
Gambar 17. Progress Bar	10
Gambar 18. Jendela "Completing Setup"	11
Gambar 19. Jendela "Xampp Control"	11
Gambar 20. Jendela "Command Prompt" untuk Setting Database	12



Gambar 21. File Library Python	12
Gambar 22. Contoh Directory Tempat File Library Python	14
Gambar 23. Copy-Paste File Extraction UG ke Htdocs	15
Gambar 24. Aplikasi Awal Ekstraksi Tabel HTML	16
Gambar 25. Contoh Tabel yang Akan Diekstrak	16
Gambar 26. Contoh Memasukkan URL dan Nama Tabel yang Disimpan	17
Gambar 27. Ekstraksi Hasil Grab Tag HTML	17
Gambar 28. Hasil Ekstraksi Tabel.....	18
Gambar 29. Database “testhtml” yang Menyimpan Tabel hasil Ekstrak	18
Gambar 30. Tabel “tesHAKI” yang Merupakan Hasil Ekstrak.....	19



USER GUIDE

APLIKASI EKSTRAKSI TABEL HTML

Aplikasi ekstraksi tabel format HTML dijalankan dengan menggunakan browser pada localhost. Aplikasi ini dibangun dengan menggunakan bahasa pemrograman Python dan menggunakan database MySQL, sehingga untuk dapat menjalankan aplikasi ini, komputer harus terinstall program Python dan database MySQL.

User Guide ini berisi persiapan dan rincian langkah-langkah yang dilakukan untuk dapat menjalankan aplikasi ekstraksi tabel HTML.

1. Spesifikasi minimum komputer untuk menjalankan aplikasi
2. Cara Instalasi Bahasa Pemrograman Python
3. Cara Instalasi Xampp yang didalamnya ada database PHPMyAdmin untuk MySQL.
4. Cara instalasi Library Python
5. Cara konfigurasi dan menjalankan aplikasi ekstraksi tabel HTML

1. Spesifikasi Komputer

Aplikasi ekstraksi tabel HTML dapat berjalan pada spesifikasi minimum komputer sebagai berikut:

- a. Processor Pentium 3
- b. RAM 512MB
- c. Hardisk 80GB
- d. Sistem operasi Windows XP/Linux
- e. Browser Internet Explorer/Firefox Mozilla/Google Chrome

2. Cara Instalasi Bahasa Pemrograman Python

Berikut ini adalah cara instalasi bahasa pemrograman Python yang berjalan pada sistem operasi Windows. Pada CD disediakan 2 jenis file installer Python 2.7.x, yaitu untuk komputer 32bit dan 64bit. Langkah instalasi Python yang berjalan pada komputer 32bit atau 64bit sama, dan berikut ini adalah langkah instalasi Python untuk 64bit

Name	Date modified	Type	Size
Extraction_UG	12/14/2013 1:17 PM	File folder	
Library_Python	12/14/2013 1:17 PM	File folder	
python-2.7.amd64	12/14/2013 11:38 ...	Windows Installer ...	15,867 KB
python-2.7	12/14/2013 1:34 PM	Windows Installer ...	15,541 KB
xampp-win32-1.8.2-3-VC9-installer	12/14/2013 11:49 ...	Application	104,004 KB

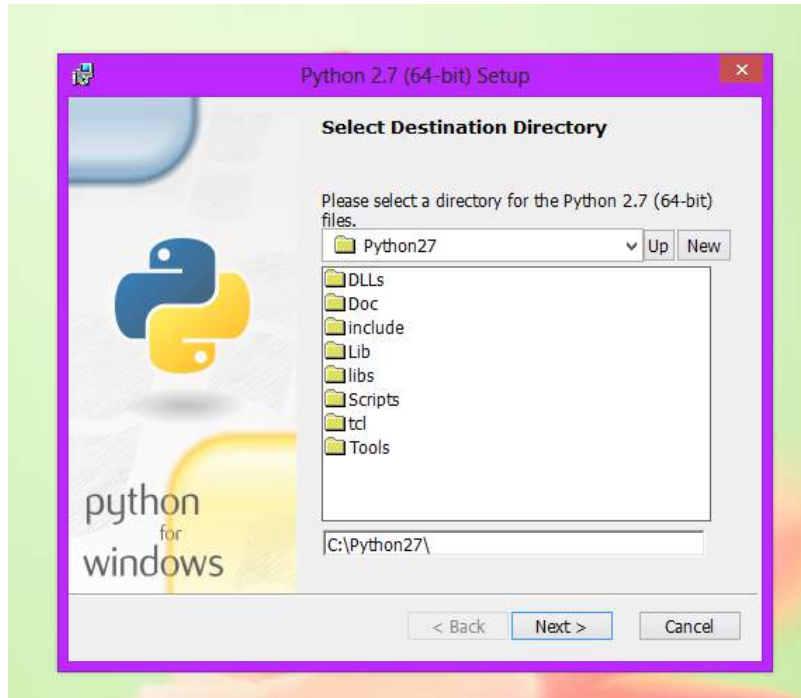
Gambar 1. File Python 2.7 Installer

Langkah-langkahnya adalah:

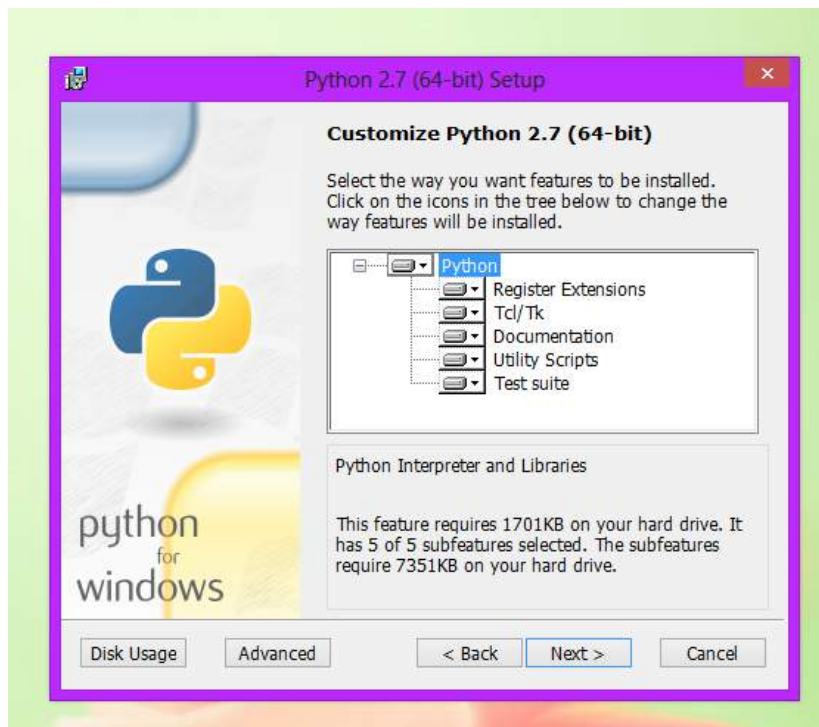
- Pilih file “python-2.7.amd64”, tampil jendela “Python for Windows”, pilih “next” (Gambar 1 dan Gambar 2)
- Pada jendela “Select Destination Directory”, pilih “next” (Gambar 3)
- Muncul jendela “Customize Python 2.7”, pilih “next” (Gambar 4)
- Muncul jendela progress bar untuk proses instalasinya, setelah itu akan keluar jendela “Completing the Python 2.7 Installer”, dan pilih “finish” (Gambar 5 dan Gambar 6)
- Maka instalasi bahasa pemrograman Python sudah selesai dilakukan.
- Setelah proses instalasi selesai dilakukan, langkah berikutnya yang dilakukan adalah melakukan “setting path”.
- “**Setting path**” dilakukan dengan membuka jendela “Properties” dengan klik kanan pada “Computer”, dan pilih “Advanced System Setting”. (Gambar 7)
- Pada jendela “System Properties” dan tab “Advanced”, pilih “Environment Variables” (Gambar 8)
- Di Jendela “Environment Variables” (Gambar 9), klik 2x variable “Path” dan jika belum ada C:\Python27 maka tambahkan ketikan berikut ini:
“;C:\Python27”, kemudian “ok” (Gambar 10)



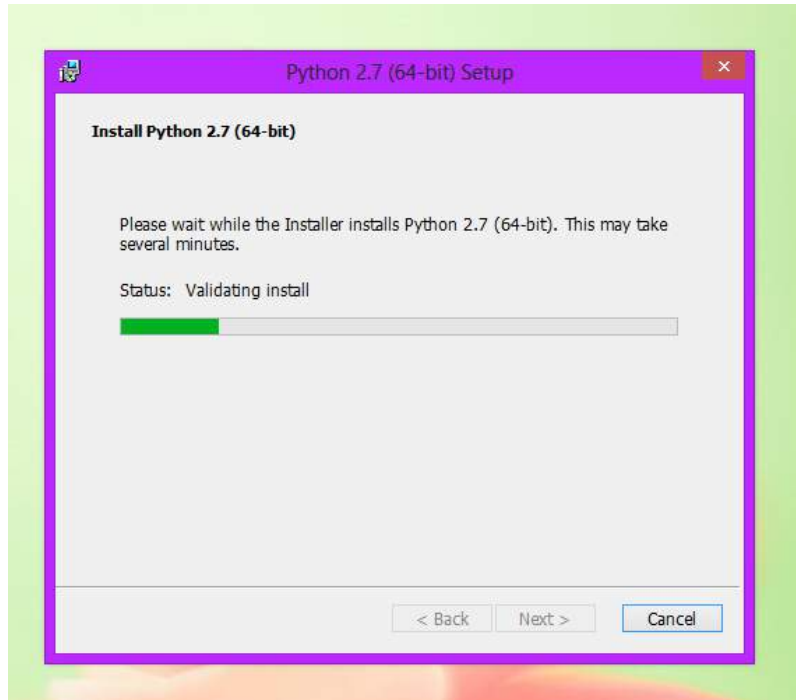
Gambar 2. Jendela Setup “Python for Windows”



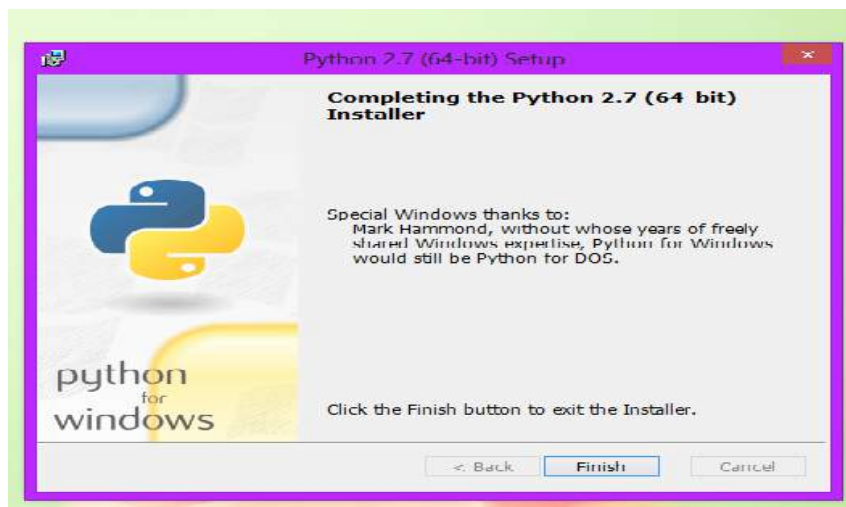
Gambar 3. Jendela “Select Destination Directory”



Gambar 4. Jendela “Customize Python 2.7”

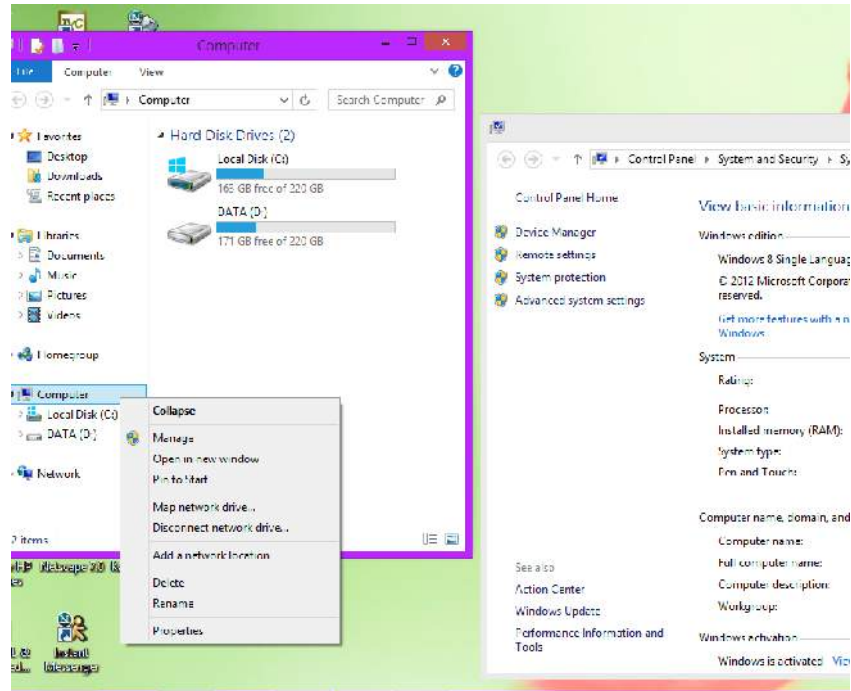


Gambar 5. Jendela Progress Bar Proses Instalasi

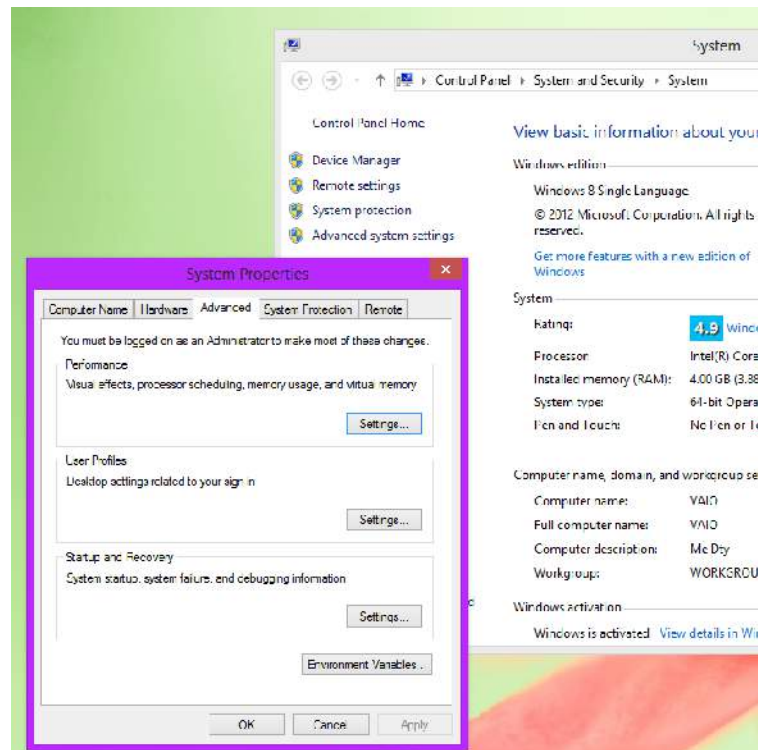


Gambar 6. Jendela “Completing the Python 2.7 Installer”

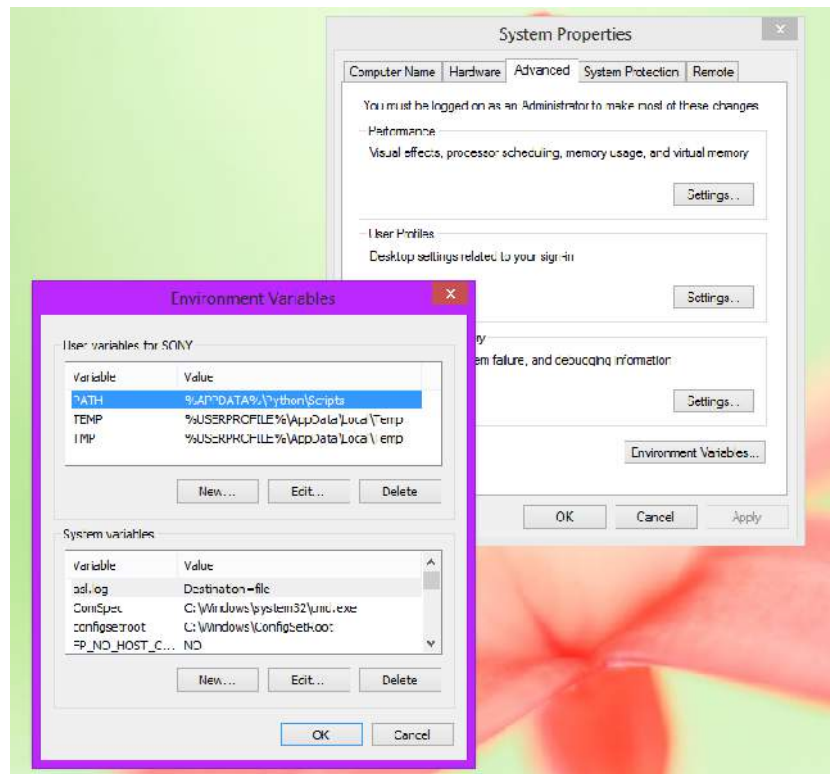
Berikut ini adalah gambar dari langkah **pengaturan path**, setelah proses instalasi Python selesai dilakukan.



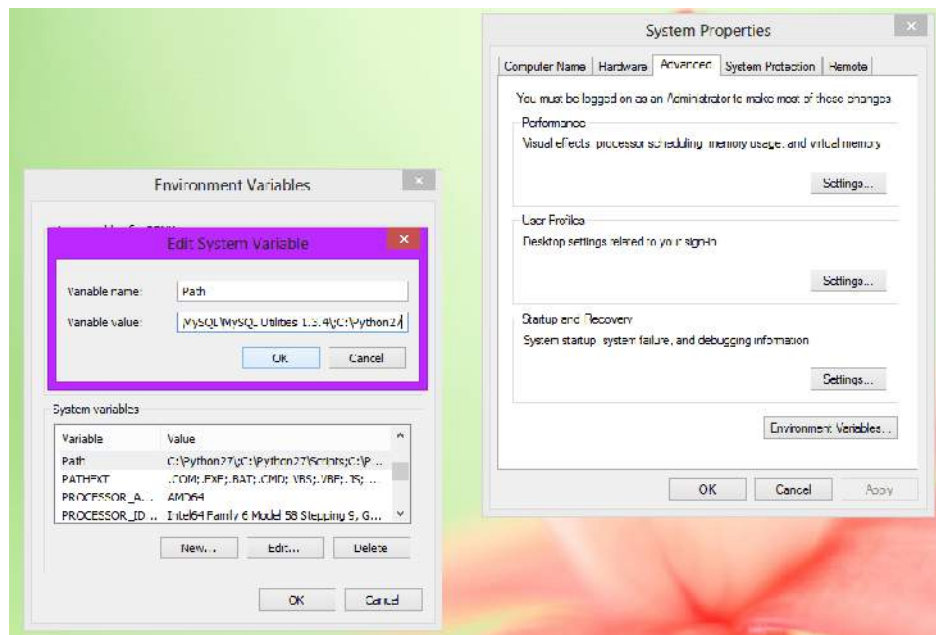
Gambar 7. Jendela “Properties” dan “Advanced System Setting”



Gambar 8. Jendela “System Properties”



Gambar 9. Jendela “Environment Variables”



Gambar 10. Jendela Variable “Path”

3. Cara Instalasi Xampp dan MySQL

Xampp adalah program yang berperan sebagai server web apache untuk simulasi pengembangan website, sehingga aplikasi web yang dikembangkan dapat diuji secara langsung dari komputer tanpa terkoneksi dengan Internet. Xampp dilengkapi dengan fitur manajemen database PHPMyAdmin seperti pada server hosting sesungguhnya. Berikut ini adalah langkah-langkah untuk instalasinya.

Name	Date modified
Extraction_UG	12/14/2013 1:17
Library_Python	12/14/2013 1:17
python-2.7.amd64	12/14/2013 11:5
python-2.7	12/14/2013 1:34
xampp-win32-1.8.2-3-VC9-installer	12/14/2013 11:4

Gambar 11. File Xampp Installer

Langkah-langkahnya adalah sebagai berikut:

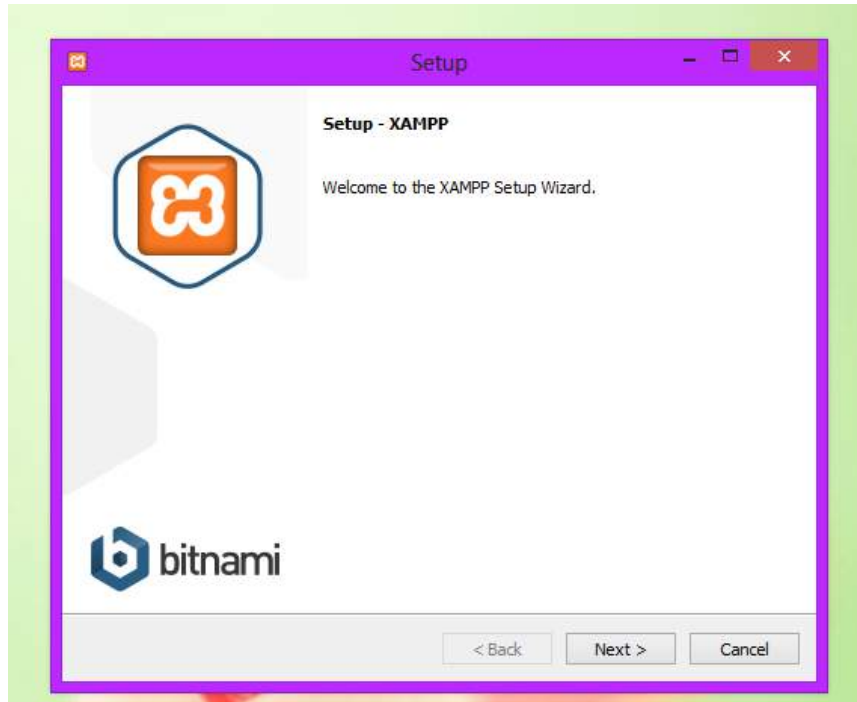
- Pilih file "xampp-win32 installer" (Gambar 11), akan tampil jendela "Set-up Xampp", pilih "next" (Gambar 12)
- Tampil jendela "Select Component", lalu pilih "next" (Gambar 13)
- Tampil jendela "Installation Folder", pilih "next" (Gambar 14)
- Tampil jendela "Bitnami for Xampp", pilih "next" dan akan keluar jendela "Ready to Install", lalu "next" (Gambar 15 dan Gambar 16)
- Tampil jendela progress bar "Set Up", jika selesai pilih "finish" (Gambar 17 dan Gambar 18)
- Instalasi Xampp dan database telah selesai dikerjakan.
- Kemudian lakukan **setting database PHPMyAdmin** pada "Command Prompt"
- Buka "Command Prompt" dan ketikkan seperti berikut:

```
C:\Users\SONY> cd\
C:\>cd Xampp
C:\Xampp>cd mysql
C:\Xampp \mysql>cd bin
C:\ Xampp \mysql \bin>mysql -u root -p
```

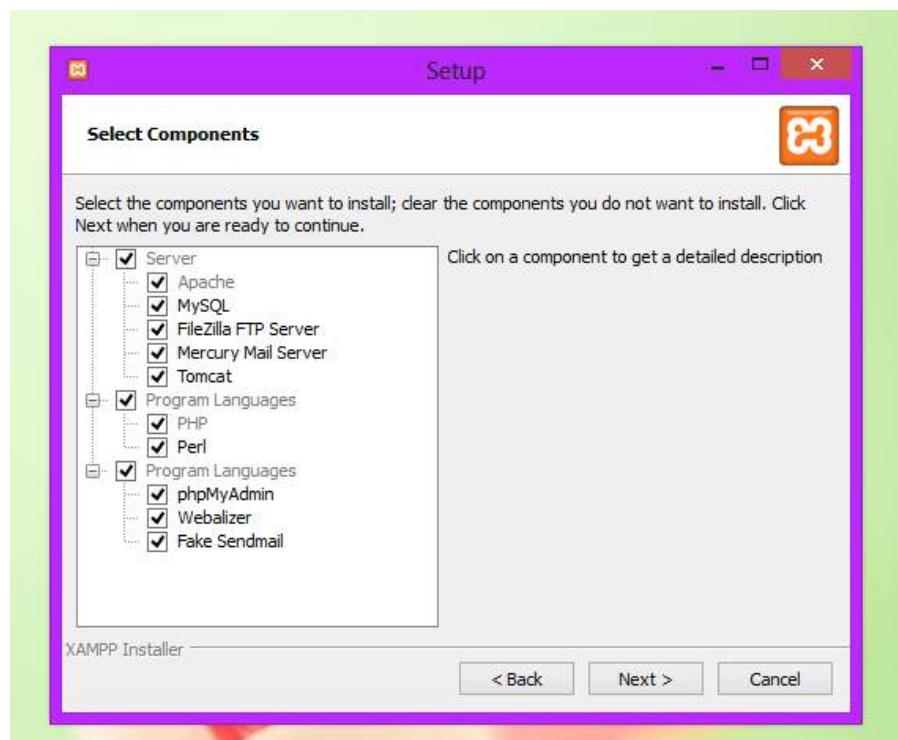
Dan karena database yang akan digunakan diberi nama: "**testhtml**", maka ketikkan pada "command prompt" (Gambar 20) tersebut sebagai berikut:

mysql>create database testhtml;

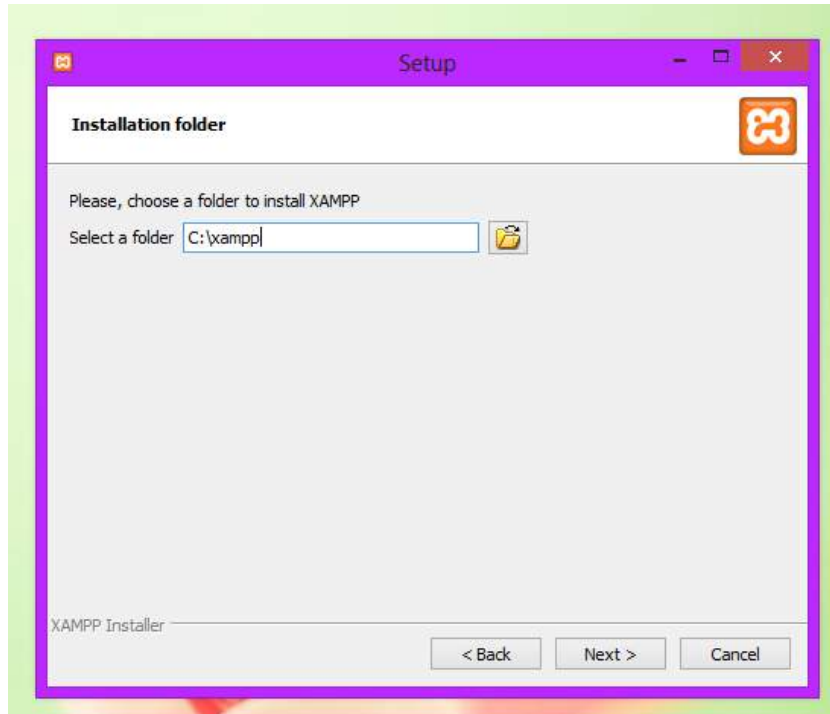
- Maka database yang akan digunakan pada aplikasi ekstraksi tabel HTML telah dibuat dengan nama: "testhtml"



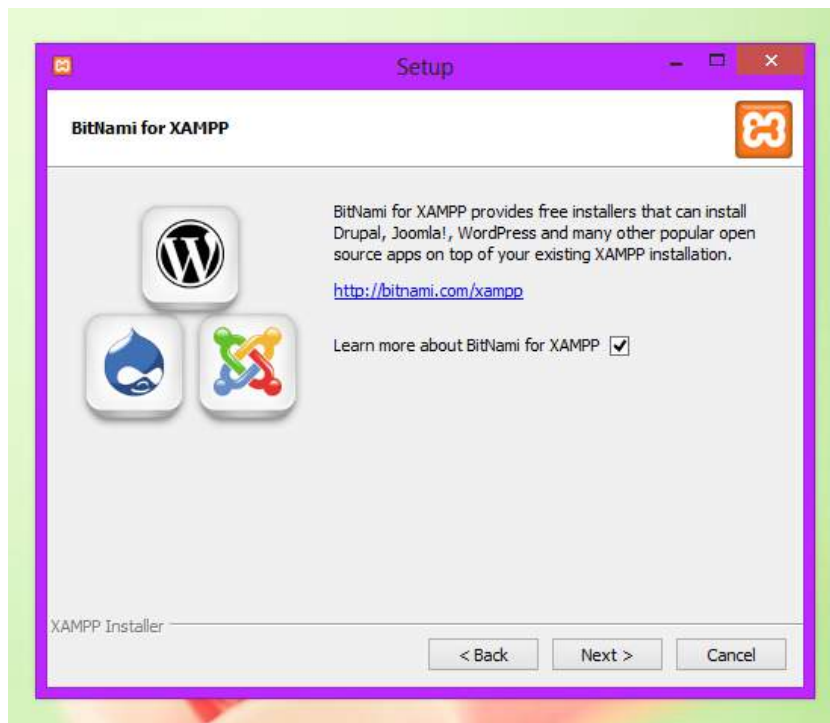
Gambar 12. Jendela “Set-up Xampp”



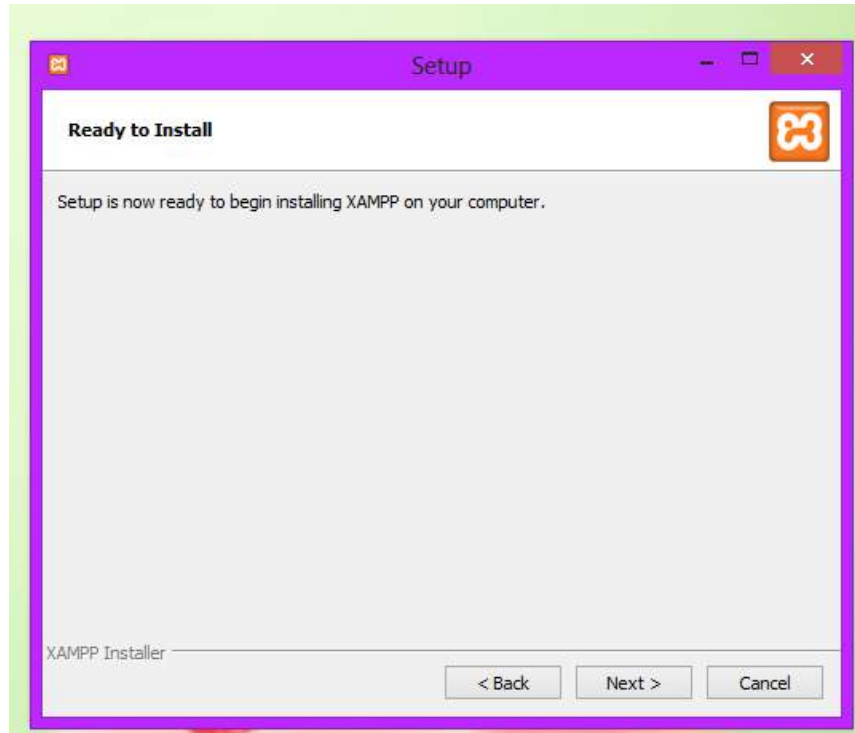
Gambar 13. Jendela “Select Component”



Gambar 14. Jendela "Installation Folder"



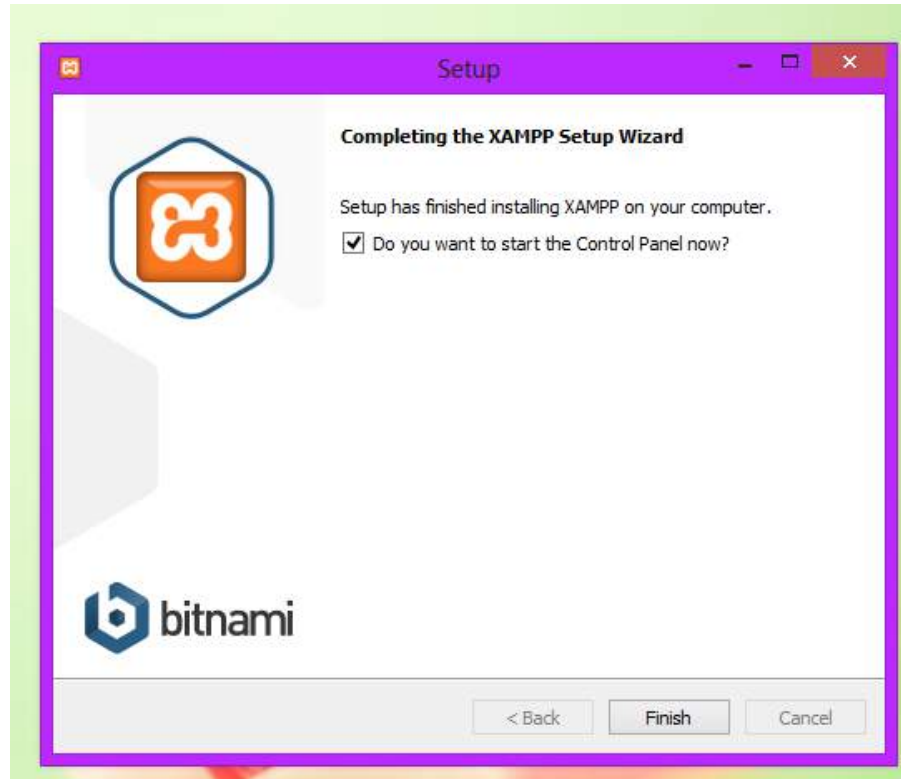
Gambar 15. Jendela "Bitnami for Xampp"



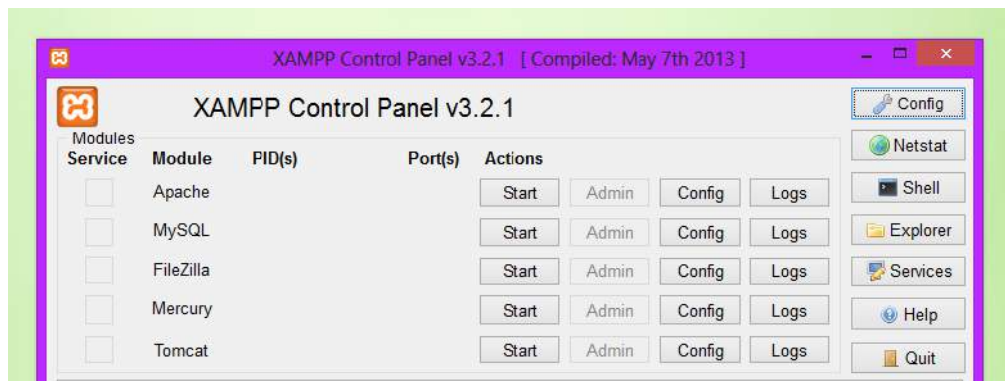
Gambar 16. Jendela "Ready to Install"



Gambar 17. Progress Bar



Gambar 18. Jendela "Completing Setup"



Gambar 19. Jendela "Xampp Control"

Pada Gambar 20 adalah jendela "Command Prompt" yang digunakan untuk melakukan setting database yang digunakan pada aplikasi ekstraksi tabel HTML.

```

Microsoft Windows [Version 6.2.9200]
(c) 2012 Microsoft Corporation. All rights reserved.

C:\Users\SONY>cd\

C:\>cd xampp

C:\xampp>cd mysql

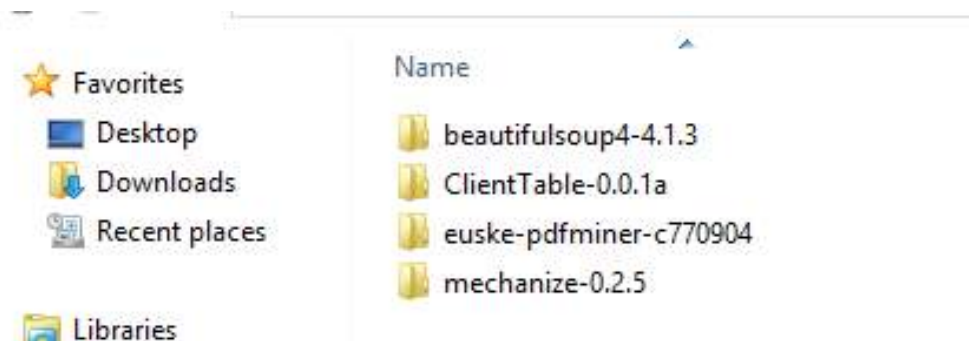
C:\xampp\mysql>cd bin

C:\xampp\mysql\bin>mysql -u root -p
Enter password:
  
```

Gambar 20. Jendela “Command Prompt” untuk Setting Database

4. Cara Instalasi Library Python

Library Python yang di install terdiri dari 4, yaitu: i). beautifulsoup, ii). ClientTable, iii). Euske-pdfminer, dan vi). mechanize. **Instalasi library python dilakukan pada “Command Prompt”.**



Gambar 21. File Library Python

Berikut ini adalah langkah-langkahnya untuk install library secara berurut.

- Buka “Command Prompt”
- Pindahlah ke directory dimana folder yang berisi kumpulan library python disimpan dalam komputer. (Gambar 22)
- Jika sudah berada di dalam directory dimana library python berada, pertama yang dilakukan adalah lakukan instalasi library untuk “ClientTable”
- Instalasi **library “ClientTable”** dilakukan dengan menambahkan ketikan pada root sebagai berikut:
 C:\.....>cd ClientTable-0.0.1a [enter]
 C:\.....\ClientTable-0.0.1a>python setup.py install [enter]

- e. Setelah instalasi library untuk “ClientTable” selesai dilakukan, ketikan:
Python [enter]
Import ClientTabel [enter]
- f. Lanjutkan dengan instalasi **library “beautifulsoup”**, dengan pindah directory terlebih dahulu, dan masuk ke dalam directory “beautifulsoup4-4.1.3”
- g. Kemudian ketikan:
C:\.....\beautifulsoup4-4.1.3>python setup.py install [enter]
- h. Setelah instalasi library untuk “beautifulsoup” selesai dilakukan, ketikan:
Python [enter]
from bs4 import BeautifulSoup [enter]
- i. Lanjutkan dengan instalasi **library “mechanize”**, dengan pindah directory terlebih dahulu, dan masuk ke dalam directory “mechanize-0.2.5”
- j. Kemudian ketikan:
C:\.....\mechanize-0.2.5>python setup.py install [enter]
- k. Setelah instalasi library untuk “mechanize” selesai dilakukan, ketikan:
Python [enter]
Import mechanize [enter]
- l. Kemudian lakukan perintah **import ulang** pada directory python, yaitu:
C:\....\Python [enter]
Import MySQLdb as mdb
Import sys
Import urllib2
Import re
Import ClientTable
From bs4 import BeautifulSoup
- m. Lanjutkan dengan instalasi **library “Euske-pdfminer”**, dengan pindah directory terlebih dahulu, dan masuk ke dalam directory “Euske-pdfminer”
- n. Kemudian ketikan:
C:\.....\euske-pdfminer-c770904>python setup.py install [enter]
- o. Setelah instalasi library untuk “Euske-pdfminer” selesai dilakukan, ketikan:
Python [enter]
Import pdfminer [enter]
- p. Library python yang dibutuhkan untuk menjalankan aplikasi ekstraksi tabel HTM telah selesai dilakukan.

```

C:\> 12/14/2013 11:49 AM 106,500,008 xampp-win32-1.8.2-3-UC9-installer.exe
      3 File(s) 138,660,776 bytes
      4 Dir(s) 184,368,558,080 bytes free

d:\DTY_PAPERS\HAKI\Python_HAKI\PROGRAM_EKSTRAKSI_TABEL>cd library_python

d:\DTY_PAPERS\HAKI\Python_HAKI\PROGRAM_EKSTRAKSI_TABEL\Library_Python>dir
Volume in drive D is DATA
Volume Serial Number is 5853-E6CC

Directory of d:\DTY_PAPERS\HAKI\Python_HAKI\PROGRAM_EKSTRAKSI_TABEL\Library_Py
thon
12/14/2013 01:17 PM <DIR> .
12/14/2013 01:17 PM <DIR> ..
12/14/2013 01:17 PM <DIR> beautifulsoup4-4.1.3
12/14/2013 01:17 PM <DIR> ClientTable-0.0.1a
12/14/2013 01:17 PM <DIR> euske-pdfminer-c770904
12/14/2013 01:17 PM <DIR> lxml-master
12/14/2013 01:17 PM <DIR> mechanize-0.2.5
      0 File(s) 0 bytes
      7 Dir(s) 184,368,558,080 bytes free

d:\DTY_PAPERS\HAKI\Python_HAKI\PROGRAM_EKSTRAKSI_TABEL\Library_Python>cd ClientT
able-0.0.1a

d:\DTY_PAPERS\HAKI\Python_HAKI\PROGRAM_EKSTRAKSI_TABEL\Library_Python\ClientTabl
e-0.0.1a>dir
Volume in drive D is DATA
Volume Serial Number is 5853-E6CC

Directory of d:\DTY_PAPERS\HAKI\Python_HAKI\PROGRAM_EKSTRAKSI_TABEL\Library_Pyt
hon\ClientTable-0.0.1a
12/14/2013 01:17 PM <DIR> .
12/14/2013 01:17 PM <DIR> ..
12/14/2013 01:17 PM <DIR> build
02/06/2003 01:46 AM 1,064 ChangeLog
02/06/2003 01:46 AM 26,437 ClientTable.py
02/06/2003 01:46 AM 1,079 COPYING
02/06/2003 01:46 AM 140 MANIFEST.in
02/06/2003 01:46 AM 525 PKG-INFO
02/06/2003 01:46 AM 4,338 README.html
02/06/2003 01:46 AM 3,756 README.txt
02/06/2003 01:46 AM 639 setup.py
02/06/2003 01:46 AM 10,282 test.py
12/14/2013 01:17 PM <DIR> testdata
      9 File(s) 48,260 bytes
      4 Dir(s) 184,368,558,080 bytes free

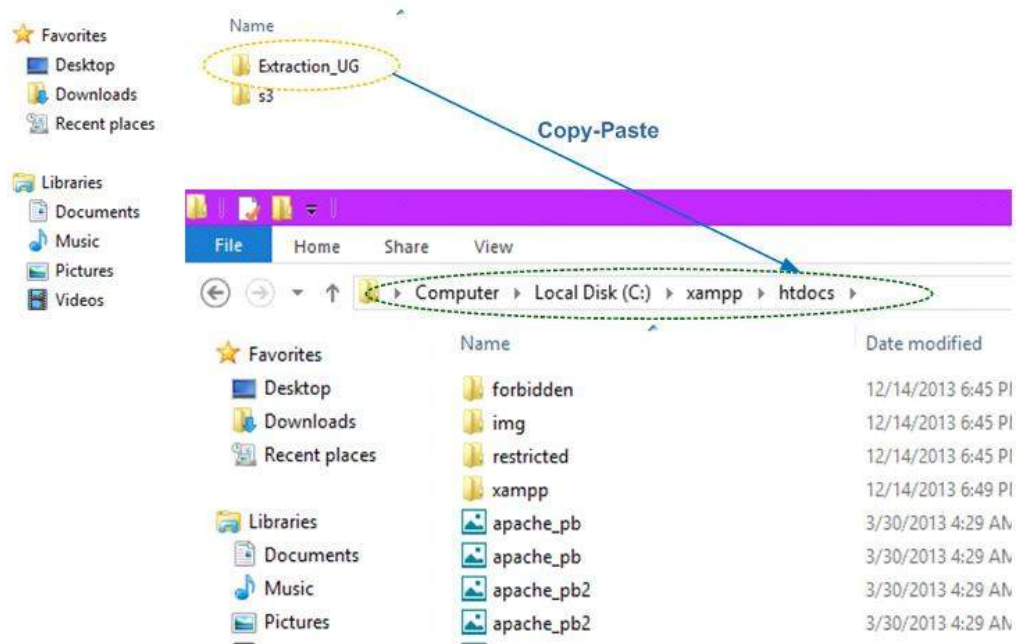
d:\DTY_PAPERS\HAKI\Python_HAKI\PROGRAM_EKSTRAKSI_TABEL\Library_Python\ClientTabl
e-0.0.1a>

```

Gambar 22. Contoh Directory Tempat File Library Python

5. Cara Konfigurasi dan Menjalankan Aplikasi EKstraksi Tabel

Program-program yang dibutuhkan untuk menjalankan aplikasi ekstraksi tabel HTML telah selesai di install. Langkah selanjutnya adalah melakukan **copy-paste** folder aplikasi ekstraksi yang diberi nama “**extraction_UG**” ke dalam folder “**C:\xampp\htdocs**”. Begitupula dengan **contoh tabel** yang akan digunakan sebagai tabel ujicoba diletakkan **didalam** folder “**C:\xampp\htdocs**”. (Gambar 23)



Gambar 23. Copy-Paste File Extraction UG ke Htdocs

Berikut ini adalah langkah yang dilakukan untuk menjalankan aplikasi ekstraksi tabel HTML:

1. Buka **“Xampp Control”** yang terdapat pada folder **“C:\Xampp”**, dan pada **“Action”** pilih **“start”** untuk module **Apache** dan module **MySQL**.
2. Buka browser, misalkan Google Chrome
3. Aplikasi ekstraksi tabel dibuka dengan memasukkan URL pada kotak address di alamat: **“localhost/extraction_UG/index.php”** (Gambar 24)
4. Pada aplikasi, pengguna diminta untuk memasukkan URL pada kotak sebelah tombol **“extract”**, URL tersebut merupakan alamat dokumen yang berisi tabel yang akan di ekstrak. Misalkan tabel yang akan diekstrak adalah tabel yang berasal dari dokumen dengan alamat: **“http://local/host/01_Ujicoba_Alg3.html”** (Gambar 25). Kemudian untuk isian nama tabel yang ingin disimpan diberi nama: **“tesHAKI”** (Gambar 26)
5. Setelah langkah 4, pilih **“extract”** dan pilih **“Ekstraksi Hasil Grab Tag HTML”** untuk menampilkan tag HTML dari dokumen yang diekstrak tersebut. (Gambar 27)
6. Pilih **“Hasil Ekstraksi”**, maka akan tampil tabel yang telah diekstrak. (Gambar 28)
7. Jika ingin mengetahui apakah tabel tersebut memang sudah berhasil diekstrak, maka pilih **“PHPMyAdmin”**. Masukkan nama databasenya, dan dapat dilihat bahwa tabel tempat menyimpan hasil ekstraksi **“tesHAKI”** sudah ada didalam database **“testhtml”**. (Gambar 29 dan Gambar 30)

localhost/extraction_UG/index.php?page=HTML



EKSTRAKSI
COLORING FUTURE

EKSTRAKSI TABEL FORMAT HTML

EKSTRAKSI TABEL FORMAT PDF

EKSTRAKSI GAMBAR

Ekstraksi Tabel Format HTML

URL: Nama file o yang ingin disimpan:

Ekstraksi Hasil Grafik Tabel HTML

Belum ekstrak

Hasil Ekstraksi

Gambar 24. Aplikasi Awal Ekstraksi Tabel HTML

file:///C:/xampp/htdocs/s3/01_Ujicoba_Alg3.html

NO	NAME			DATE OF					
	FIRST	MIDDLE	LAST	BIRTH			DEATH		
				D	M	Y	D	M	Y
1	Angga	Kamandanu	Putra	01	05	91	n	n	n
2	Bian	Rangga	Aditya	02		92	n	n	n
3	Cika	Mutiara	Anisa	03	06	93	n	n	n
4	Dewi	Anggara	Putri	04	07		n	n	n

Gambar 25. Contoh Tabel yang Akan Diekstrak

Gambar 26. Contoh Memasukkan URL dan Nama Tabel yang Disimpan

Gambar 27. Ekstraksi Hasil Grab Tag HTML

ost/extracton_UG/index.php?page=HTML&k=2

NO	Alamat	NAME FIRST	NAME MIDDLE	DATEOF BIRTH D	DATEOF BIRTH M
1	Dekasi	Angga	Kamandana	01	05
2	Takala	Bian	Rangga	02	02
3	Begor	Cika	Mitara	03	06
4	Tepok	Dewi	Anggara	04	07

Gambar 28. Hasil Ekstraksi Tabel

Gambar 29. Membuka Database “testhtml” yang Menyimpan Tabel hasil Ekstrak

Ekstraksi Tabel Format HTML

URL:
Extract **Clear**
Nama tabel yang ingin disimpan:

Ekstraksi Hasil Grab Tag HTML

Hasil Ekstraksi

PHP MyAdmin

- ☐ tabel236
- ☐ tabel241
- ☐ tabel242
- ☐ tabel243
- ☐ tabel243coba
- ☐ tabel244
- ☐ tabel245
- ☐ tabel246
- ☐ tabelcoba
- ☐ tabeljoincol1
- ☐ tabeljoincol2
- ☐ tabeljoincol3
- ☐ tabeljoinrow1
- ☐ tabeljoinrow2
- ☐ tabeljoinrow3
- ☐ tabelsederhana1
- ☐ tabelsederhana2
- ☐ tabelsederhana3
- ☐ tabelsederhana4
- ☒ teshaki
- ☐ test123
- ☐ testdty
- ☐ testlu

[[Create PHP
de]]
[[Refresh]]

w #

mode and

NO	Alamat	NAME_FIRST	NAME_MIDDLE	DATEOF_BIRTH
1	Bekasi	Angga	Kamandanu	01
2	Jakarta	Bian	Rangga	02
3	Bogor	Cika	Mutiara	03
4	Depok	Dewi	Anggara	04

w #

mode and

Gambar 30. Tabel “tesHAKI” yang Merupakan Hasil Ekstrak